Available online at **www.joac.info**

ISSN: 2278-1862

Journal of Applicable Chemistry

2023, 12 (2): 92-113

(International Peer Reviewed Journal)

Artificial Intelligence-2 Tools in Modelling Chemical Response[#]: Part III. Neural Network Modelling and Prediction of ^{13}C NMR Response of Halomethanes

K. Somasekhara Rao[1], K. Ramadevi[2], K. Ramakrishna[3], Ch. V. Kameswara Rao[3], K. M. M. Krishna Prasad[4] and R. Sambasiva Rao[4]*

1. Dept. of Chemistry, Acharya Nagarjuna Univ., Dr. M.R.Appa Rao Campus, Nuzvid-521 201, **INDIA**
2. PG Department of Chemistry, Sir C.R. Reddy College for Women, Eluru-534001, **INDIA**
3. Department of Chemistry, Gitam Institute of Science, Gitam University, Visakhapatnam-530 017, **INDIA**
4. Department of Chemistry, Andhra University, Visakhapatnam 530 003, **INDIA**

Email: sraokaza1947@gmail.com, karipeddirk@gmail.com, rsr.chem@gmail.com

(Dedicated to I. Suryanarayana, IICT, Hyderabad during his 70th birth anniversary)

Accepted on 16th March, 2023

ABSTRACT

Neural networks (NNs), a data driven second generation artificial intelligence technology is employed to model and predict ^{13}C NMR chemical shifts of 23 halomethanes with charge and functions of polarizability (α and α^2). Optimum single layer perceptron (SLP), 4-layer multi layer perceptron (MLP), radial basis function (RBF) and generalized regression (GR) neural network (NN) models (i.e. architectures, transfer function) are reported using the commercial software package TRAJAN 5.0. The analysis of data sets with (NP = 23) and without outliers (NP=18) using explanatory (causative) variables (α, C and α^2, C) was performed with IPS (Intelligent Problem Solver), a fast solution provider of TRAJAN and designed experimental runs for the task. The optimum set of models adequately explaining and predicting ^{13}C NMR response are SLP 2-9-1, 2-5-1; MLP 2-6-2-1, 2-5-5-1; RBF 2-16-1 and GRNN 2-23-2-1 (with smoothing factor in the range of 0.01 to 0.02). The best models of the best set based on the principle of parsimony are SLP 2-9-1 for full data set and SLP 2-5-1 for a subset of 18 compounds. The results of test data sets ensure the single and two compound predictions are within tolerable error limits. Multiple linear regression (MLR), a hard regression model, least median squares (LMS) and partial least squares regression (PLSR) (soft) models are inadequate even in explaining the variance in response, leaving the only alternative of non-linear input-to-output (I/O) mapping. The simulated and ^{13}C response data sets belong to function approximation task and the imbibing NN paradigm is not an alternate but indispensable tool for the current endeavour.

Keywords: ^{13}C NMR, Halomethanes, Neural Networks, PLSR, Artificial intelligence, Function approximation.

Part I : Ann Chim Rome 95(2005) 291-301; Part II: Asian J Chem, 22(8)(2010)5937-5950

INTRODUCTION

Chlorofluorohydrocarbons (Freons) have been mostly replaced with alternate moieties with similar properties in refrigeration industry [**1**] . Due to the high stability of freons which escape in to the stratosphere, their concentration increases over time. It is one of the factors for the depletion of ozone layer. The consequent increased passage of UV light results in health hazards and environmental issues. This brought renaissance in probing into studies [**2**] of structure-property (SPropR), structure-activity (SActR) and structure-response relationships (SRespR).

$^{!3}$C NMR spectroscopy is one of the versatile tools in quantifying the structure of the compounds. The very same technique was popularized as magnetic resonance in medical sciences [**3-6**]. It is with a positive motto to drive away of the psychological stigma of ill effects of the nuclear phenomenon. Experimental NMR spectral profiles, CQC (computational quantum chemistry) computed values, chemometrically predicted data are in backend in organic compound structural elucidation, SXR (Structure X relationships) and probing into proteins, soft tissues as well as nano materials. Apart from voluminous ^{1}H NMR spectral data [**7**], interest in ^{15}N [**8**], ^{19}F [**9-10**], ^{31}P [**11**], ^{27}Al [**11**], ^{51}V [**12**] etc. emphasizes the prospects of multi nuclear NMR. Functional MR imaging (fMRI) in 3D- (three dimensional) and 11.7 Tesla-MRI revolutionized detection of ischemic heart disease [**13,14**], migraine [**3**], muscular skeletal defects [**4**]and detection of metabolites [**5**] in children with brain tumour. The chemical shifts in ^{1}H NMR are applied to predict gasoline density [**15,16**], boiling point, bio-diesel [**17**], quality of cheese [**18**] employing chemometric data processing algorithms like partial least squares (PLS), hierarchical cluster analysis (HCA), discriminate analysis (Linear DA, Quadratic DA) and NNs (MLP, RBF, LVQ, SOM, neural gas etc.). For human imaging, 15 to 20 T-MRI are under active investigations in select research institutes [**4**]. The experimental spectral databases of ^{13}C-NMR [**19, 20**] for a large number of compounds are now available in hard, soft forms as well as on web. ^{13}C NMR shifts are the basis in predicting classification of disaccharides/tri-saccharides [**21**], polymers [**22**]and human ubiquitin. Random forest (RF), CART (Classification And Regression Tree), PLS (partial least squares) etc. were the hard/soft modelling approaches to arrive at the solution. The prediction of diffusion co-efficient [**23**] and vapor liquid equilibria [**24**] was reported using one energy parameter. Advanced experimental procedures, hyphenated instrumental techniques, E-man, Chemometrics, AI2 and explainable AI along with automatic rule/knowledge extraction gadgets, Consciousness tools, mimics of nature's intelligence in life/non-life forms are now in vogue to detect a single molecule and atto-gram materials and to prepare tailor made compounds/materials for medicine/ industry/ comfort/ aeronautics.

No theoretical model from first principles is available for variation of bio-, physico-and/or chemical properties/responses in Pharmaceutical, Biological and Environmental chemistries. Neural networks, data driven models, imbibing non-linearities with a few basis functions have surpassed empirical and non-linear least squares in obtaining end results. The selection of relevant descriptors improves not only the success rate, but also diminishes the false positive and false negative conclusions. In mega projects [**25**], a large number of compounds, multiple responses, big pool of descriptors, host of pre-processing and transformation techniques have been used to arrive at the best set of models. However, the modelling activity at the moment appears to be a craft as there is no set path for different phases and exhaustive modelling is impracticable **[2**]. Braibanti *et.al* [**26**] reported multivariate calibration with ion selective electrodes (ISEs) using RBF and fuzzy-ARTMAP neural networks. Rao *et.al* [**27**] proposed data driven NN models in process chemistry of ceramics, variation of rate constants with hydrogen ion/GC response of organo phosphorous compounds and classification of ores/minerals [**27**]. Gasteiger and Suryanarayana [**28**] reported least squares analysis of ^{13}C chemical shifts for subset of twenty halomethanes excluding some of fluoro derivatives. Clark et.al [**29**] employed atomic charges (monopole/ dipole/quadrupole) and bond order to explain variation of ^{13}C NMR of a set of 45 organic compounds employing MLP-NN with back propagation training algorithm. Yoshida et al [**1**] found NN [**30, 31**] to be better than PLS and QPLS for the variation of boiling points of 48 halomethanes. Hervai [**32**] simulated ^{13}C NMR spectra for 15 lignins

with geometric and electronic descriptors calculated from 3D-optimized structure at AM1 level and topological ones from 2D-representation of molecules.

We report here the modelling and prediction of ^{13}C chemical shifts of 23 halomethanes with MLP, RBF and GRNN architectures.

MATERIALS AND METHODS

NMR Data: The data set consists of ^{13}C NMR shift of methane and 23 halo substituted ones. The explanatory variables considered are charge and α and data matrix (x) is of size 23 x 2, while for response (y) is 23 x 1. Thus, it is a 2-way data set of multivariate and uni response system experimentally obtained from NMR in zero order mode.

Hardware and Software: DELL Inspiron/6000 model with 504 MB RAM with a CPU of clock speed 1.86 GHz and a hard disk of 74.4 GB. MVATOB (MultiVariate Analysis Tool Box) in MATLAB [33] from Math work is employed both for statistical and visual exploratory data analysis (SEDA and VEDA). MVATOB developed in our laboratory consists of a suite of object oriented m(atlab function) (Om) files ($$$.m) for least squares (LS), least median squares (LMS), principle component regression (PCR), partial least squares regression (PLSR) modelling and visualisation (graphics, surfaces and contours) of the input and output. Trajan 5.0 [34] is employed for neural network modelling.

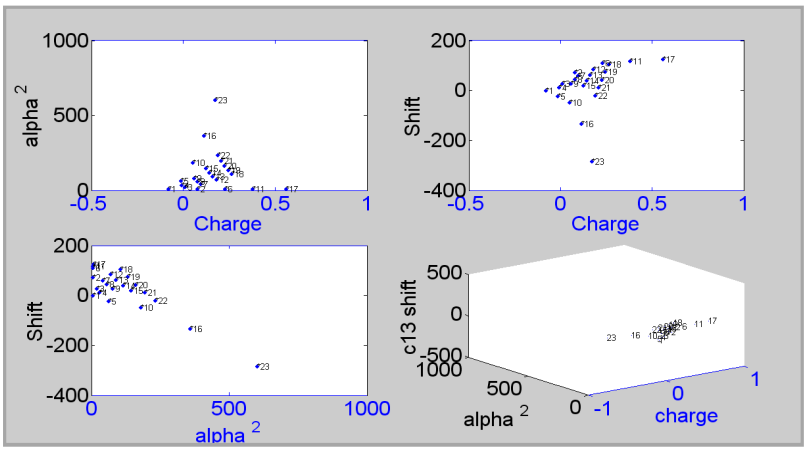
Exploratory analysis (SEDA, VEDA): Profiles and Correlation between Chemical Shift and Predictor variables. The predictor space consisting of α, charge (C) and their functions to model the change in ^{13}C NMR chemical shifts of a set of 23 halomethanes (Table 1) is explored.

Table 1. ^{13}C NMR response, charge and polarizability of Halomethanes

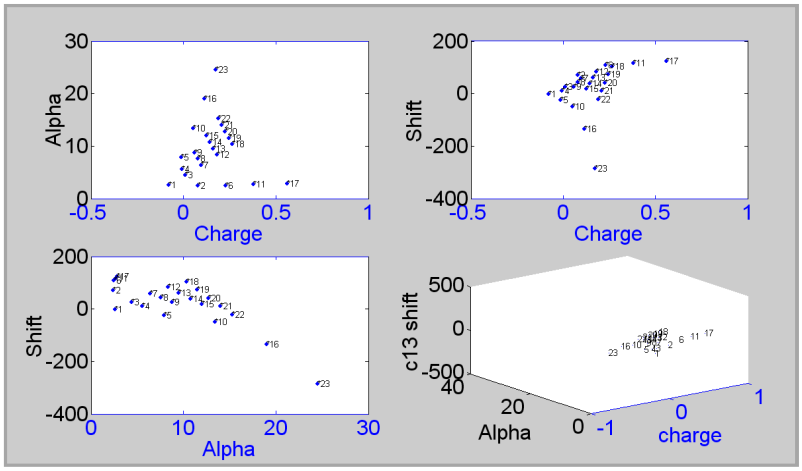
No	Compound	Charge	Alpha	^{13}C Response
1	CH4	-0.078	2.601	-2.3
2	CH3F	0.079	2.394	71.57
3	CH3Cl	0.011	4.43	25.56
4	CH3Br	-0.008	5.549	9.64
5	CH3I	-0.012	7.886	-24.04
6	CH2F2	0.23	2.479	109
7	CH2Cl2	0.097	6.376	58.3
8	CH2ClBr	0.078	7.532	43.4
9	CH2Br2	0.059	8.768	26.3
10	CH2I2	0.052	13.427	-50
11	CHF3	0.38	2.65	116
12	CHCl3	0.181	8.342	82.3
13	CHCl2Br	0.162	9.514	60.7
14	CHClBr2	0.143	10.747	38
15	CH2Br3	0.125	12.011	17.3
16	CHI3	0.114	18.983	-135
17	CF4	0.561	2.858	122.3
18	CCl4	0.266	10.315	102.3
19	CCl3Br	0.246	11.497	72.9
20	CCl2Br2	0.226	12.727	40
21	CClBr3	0.207	13.985	10
22	CBr4	0.189	15.262	-22.7
23	CI4	0.174	24.544	-287

A perusal of the scatter plot (Figure l(a)) shows a set of straight lines with different slopes and intercepts. Pearson linear correlation coefficient of α and charge is –0.02 indicating that they are statistically not correlated. Further, the angle between the two (α and charge vectors) is 50^{o} inferring

that they are non-orthogonal. The uni–, multi–variate (3D response surface and 2D-contour) ^{13}C NMR chemical shift against α and charge are given in figure1.



(a). NMR01.dat



(b) NMR11.dat

Figure 1. 2D- and 3D- Scatter plots of C13 NMR response with Charge and (a) alpha^2 (b) alpha.

Since the variation of the chemical shift with α is not linear, higher powers of α are attempted. Since the partial correlation of chemical shift with α, α^2, α^3 and charge are -0.80, 0.89 and –0.91 respectively, the improvement from α^2 to α^3 is not significant. Although α and α^2 can be considered as explanatory variables, α^2 alone is preferred based on the correlation coefficient of 0.95 between α and α^2. A similar study with charge shows that cc of shift with c, c^2 and c^3 are 0.34, 0.36 and 0.33 and cc between C and c^2 is –0.90 indicating that c is adequate.

The percent explainability of variance in chemical shift and ratio (F) of sum of squares of residuals (SSR) to total sum of squares (TSS) with regressors are considerably higher for (α^2, C) compared to (α, C) (Table 2). Thus the model (1) is to be preferred over (2).

$$\text{Shift} = a_0 + a1 * \alpha^2 + a2 * \text{charge} \quad \ldots (1)$$
$$\text{Shift} = b_0 + b1 * \alpha + b2 * \text{charge} \quad \ldots (2)$$

(Hard) Least Squares and (Robust) Least median squares procedures: The results of unit weighted multi linear (ML) LS and ML-LMS regression analysis for models (1) and (2) are presented in table 2 along with the expert system inferences. The variance scaled residuals for the regression model 1 is depicted in Supl-figure 1. ML-LMS detected the data points 6, 11, 17 and 23 as outliers based on scaled residuals and MAD statistic [**35, 36**]. However, ML-LS could not detect any one of

Table 2. Expert system inferences of Correlation, ANOVA and MLR

	Alpha^2 , Charge				Alpha , Charge			
	NMR01	NMR02	NMR03	NMR04	NMR11	NMR12	NMR13	NMR14
Data points deleted	--	6,11,17	6,11,17,23	6,11,17, 23,5	--	6,11,17	6,11,17,23	6,11,17,23,5
% Explainability	91.07	99.43	98.89	99.36	73.54	90.04	92.88	92.58
Inference	.T.	.T.	.T.	.T.	.T.	.T.	.T.	.T.
F value	101.96	1483.12	715.15	1156.44	27.80	76.80	104.33	93.52
Validity of parameters	.F.	.F.	.F.	.F.	.F.	.F.	.F.	.F.
CC(alpha, charge)	0.02	0.40	0.47	0.45	-0.02	0.54	0.58	0.58
Inference	Valid	Invalid	Invalid	Invalid	Valid	Invalid	Invalid	Invalid
CC(Response, alpha)	-0.88	-0.86	-0.64	-0.68	-0.78	-0.73	-0.50	-0.54
CC(Response, charge)	0.36	0.11	0.38	0.34	0.36	0.11	0.38	0.34
Inference	Valid	Valid	Invalid	Invalid	Valid	Valid	Invalid	Invalid
ES advice	Invalid	Invalid	Invalid	Invalid	Invalid	Invalid	Invalid	Invalid
	NMR01	NMR02	NMR03	NMR04	NMR11	NMR12	NMR13	NMR14

the outliers as all the residuals are within the range of –3.0 to +3.0, in spite of the fact the magnitude of res-LS are significantly greater than the precision of response. Thus it is an instance that ML-LS method fails in detecting the outliers. Hence, this model does not explain the variation of the chemical shift. The outliers (6, 11 and l7) are removed and the data set NMR12.dat is analysed (Supl-Table 2). The points 23 and 5 exhibited high values of scaled LMS residuals and sdy decreased from 46.71 to 6.48. The cropping up of these outliers after elimination of first set (viz. 6, 11 and l7) is an instance masking effect.

The regression analysis after eliminating the data point 23, the file NMRl3.dat still indicated the presence of an influential outlier (data point 5). The results of the data file (NMR l4.dat) devoid of outliers 6, 11, l7, 23 and 5 show that the regression coefficients of LS (sensitive to outliers) and LMS (robust to outliers) are comparable. The scaled residual plots infer that all the residuals are within –l.5 to 2.0, inferring that there is no need of further LMS iteration cycle. The significant F for regression and low sdy endorses the validity of this statistically adequate model for the data set NMRl4.dat. The model is invalid based on t-test for regression coefficients. Further the model parameters are unreliable, since their standard deviation exceeds 33% of the parameter values. Hence, the question of prediction does not arise at all. The magnitude of regression coefficients by LMS and LS are significantly different which is an indicator of the presence of outliers. The ML-LMS and ML-LLS iterative cycle is performed for model2, wherein the explanatory variables are α and C. Similar trend is observed except that the residuals are much higher compared to the model with α^2 and C.

Soft regression procedures: PCR and PLSR: The very low correlation coefficient (-0.02) between α and C results in principal component (PC) axes coinciding with those of explanatory variables axes. The results of PCR are almost identical with MLR as expected. The covariance of Y (shift) with X (α^2, C) is significant and partial least squares (PLS) analysis is carried out. The two PLSCs explained 99.9 and 0.1% of covariance of the data set (NMR11.dat) respectively. The vectors are orthogonal. The results of polynomial (of order 1, 2 and 3) PLSR analysis with the successive removal of outliers are depicted in (Supl-Table 3).

The scatter profiles of shift versus the two PLSC vectors for model 1 in 2D- and 3D- space and the substantial decrease in sdy (from 40.31 to 25.97) with the introduction of second PLSC vector, suggests the necessity of both the components in regression. However, the SD in regression coefficient (Supl-Table 3) corresponding to second PLSC in the regression equations invalidates the two-component model for further use of parameters. However, it explains the variance in y for the compounds under investigation.

Shift = [PLSCl PLSC2] * [a-PLSC1 a-PLSC2]

PLSR is repeated for the data files NMR11.dat, NMRl2.dat, NMRl3.dat and NMR14.dat, which are generated by successive elimination of outliers as discussed earlier. A perusal of Supl-Table 3 reveals that sdy decreases and percent explainability increases monotonously. The validity of PLS regression parameters and the application of the models in explaining the variation of shift are given in the last two rows.

The scatter plot of scaled residuals in ^{13}C NMR shift with one- and two– PLS components clearly depicts that the residuals are randomly distributed (no trend for a consecutive set of data points). Further two-component linear-PLSR is adequate while one component model is inadequate. The second order (with squared and cross product terms) and cubic (linear, quadratic and cubic terms) PLSR models (Supl-Table 3) are attempted. The decrease in sdy from linear to quadratic models is significant, while for cubic PLSR it is not substantial. From curve fitting framework, quad-PLSR is a better model. But the SDs in parameters corresponding to quadratic and cross product terms invalidate the t-test for the regression coefficients of the model. The cubic PLSR model is invalid as well as over ambitious.

All PLSR models are studied for model 2 with explanatory variables α and C. But for the fact that the residuals are much larger, the trend in regression coefficient and sdys are similar. The number of parameters increases with the order of polynomial in PCR and PLSR procedures. The large SDs in regression parameters indicates small number of data points corresponding to the function of explanatory variables in the model.

The data set NMRl4.dat is devoid of statistically detectable outliers. But the unscaled residuals in chemical shift for some of the compounds are more than 20%. It indicates non-linearity in the functional relationship between the response and causative variables, inappropriate/less influencing missing explanatory variables or gross error in response. Since the magnitude of unscaled residuals for the bromo compounds (3, 4, 9, 15 and 21) are more than the precision accuracy of the experimental shifts, they are eliminated in the data set NMR15.dat. The results of both LLS and LMS are far superior to statistically valid model, NMR14.dat. A close inspection of results of hard and soft modelling procedures and multi dimensional scatter plots in explanatory as well as response space vividly show several distinct linear clusters along with singletons.

The entire data (NMR11.dat) may be viewed as sets comprising of different halo (flouro, bromo, iodo and mixed) or hydrogen atoms (CH3X, CH2XY, CHXmYm, CXmYm) in halomethanes. The results of the regression analysis of all possible subsets of data confirm only 3 to 4 compounds follow a linear model. The different linear models cannot be generalised even by piecewise or by a polynomial regression procedure. Thus, neural network, a data driven non-linear I/O mapping with excellent function approximation characteristics is adopted.

E-man (Evolution of Mimics of Algorithms of Nature): E-man [27] is a set of software implemented algorithms formulated with the inspiration from foraging, building/shifting of ant colonies/honeybee hive, migrating birds/fish schools, genetic inheritance/diversity and nervous system in animate world and brain/mind/consciousness/immune-system of human beings. Each category found a separate niche as a paradigm and evolving from the standpoint of mapping of scientists' perception of nature.

However, E-man (even with explainable AI) is not a panacea but yet another option in information processing. Three decades of concerted efforts unequivocally proved their excellence in solving classification, function approximation, and multiple constrained/multi object optimization tasks in science/engineering compared to popular theoretically sound inductive/deductive and gradient based approaches. In spite of the fact that, nature mimicking algorithms are invariably outcome of partial understanding of intricate complicated processes, they are instrumental to probe into how biological systems construct their representation spaces. The blue brain is an artificial rat

brain partially i.e. a cellular level computer system just mimicking a two week old rat somato sensory neo-cortex. The blue chip of chess and humanoid robots not only brought laurels to AI, but lay a promise for future hyper intelligence systems. This new era started with McCulloch and Pitts' artificial neuron in 1940s which exploded into ANNs, the future Connectomics, the hope of artificial human brain. Symbolic expert systems, mechanical robots, generalized problem solver and theorem proving approaches were nurtured in 1960s under the umbrella of first-generation AI products. Numerical ESs, knowledge extraction methods, E-man and humanoid robots are the tools of AI2. The man-machine-interface is on the way to realize artificial brain with an intermediate sub goal to understand human brain and/or to intervene with disorders/probe into the mind/consciousness. Electronic-nose, E-tongue, E-eye etc. are now prototype products which emerged as a result of integration of electronics, chemical molecules and information science embedding omni-metrics [**27**]. The environmental/ medicinal/ industrial chemistries are also reaping the benefits of cross fertilization of nature mimicking approaches in hierarchical information generation/processing. This paves way to generate super/hyper intelligent interfaces and algorithms to emulate/ reach nature as close as possible. QC is a computational chemistry approach based on solving a second order ODEs (ordinary differential equations), the heart of Schrödinger wave equation. This paradigm attained the experimental accuracy through hybrid approaches. Recent advances in computational chemistry exploited NNs in QC advantageously. NNs deal with data driven robust models using primary/derived experimental data/bio-physico-chemical parameters. The prediction of ^{13}C NMR has become robust and more accurate in this decade with data driven models and nature inspired approaches.

Data driven modelling paradigm with NN: Neural Network (NN) models pervade in all branches of science and sky is the limit for applications. Function approximation, classification, character recognition, pattern classification/recognition are a few of the common tasks solved with NNs. The imbibing character and versatility of NNs is now a reality in emulating well-established mathematical/statistical procedures. In this decade neural networks are successfully developed for non-linear correlation analysis [**35-37**], finite element method [**38**], numerical solution of ODE [**39**], Runge Kutta method [**40**], Lorenzian and rational functions [**41**], convex programming [**42**], calculation of inverse Jacobian [**43**], and solution of linear/ quadratic equation [**44**] and Boolean functions [**45**]. Further the nonparametric hypothesis testing is emulated with genetic algorithm inside NN [**46**].

We reported feed forward (FF) NN architecture to estimate rate constants for integral order kinetics without a priori model [**47**]. Gemperline [**48**] reported a full quadratic model for calibration of active constituent in pharmaceutical preparations with MLPs. Kowalski [**49**] was the first to propose Chemnets where the unexplained response (after the absorbance vs. concentration data was fitted with a linear model) was modelled with quadratic function. The stability constants for MLH species and variation of rate constants were estimated with a new chemical theoretical net in our laboratories [**50**].

Function approximation: Mathematically universal approximation theory was proved for any shape, of course restricted to certain class of basis (shapes) functions. In NN paradigm, mathematical function approximation is proved with Sigmoid and Gaussian transfer functions (TFs). But recently several other basis functions are also reported. A few typical ones are pseudo Gaussian [**51**], raised cosine [**52**], ridge [**53**], binary cube [**54**], multi-wavelet [**55**] with ortho normal multi scaling, Tikhonov regularization [**56**] and clustering functions. The multivariate functions [**57**], (those with high input variability), smooth functions [**58**] etc., are considered in function approximation. In addition to normal mathematical space, Sobolov space is also considered [**59**]. Clustering for function approximation was used to initialise centres of RBF [**60**].

However, universal approximation theorem does not explicitly tell the number of neurons, layers, TFs, tools for solution for a specific problem. The number of parameters, their characteristics makes the solution of a problem hard and NL complete. Hence, it is solved with various approximations at

different phases. Respecting the constraints imposed in universal approximation theory, it is possible to arrive at a RBF or MLP-NN (with number of neurons less than the number of points), which reproduces the original shape (1-D to m-D).

Yet, the hurdles are noise, outliers, and correlation in data structure to obtain a solution of desired quality. The optimisation method, object function (performance, error, parameter reliability) and ranges of data play a major role in solution, generalizability, reliability, predictability etc.

Radial basis function (RBF) neural network (NN): RBFnet [31] is another sought after competing NN with MLP, a basic unit in the hands of neural computational scientists. A three-layer RBF with hidden layer being radial layer and four-layer RBF containing a sigmoid TF in the second hidden layer are proposed to account for Gaussian alone or a product of Gaussian and sigmoid functions as basic unit to model complex multi-dimensional space. Trajan implements 3-layer RBF employing pseudo inverse to train the data set. Professional II has a provision to invoke both RBF models and training can also be affected by all other popular ones in NN viz. back propagation (BP), delta-bar-delta (DBD) etc. The variable parameters in RBF implementable in TRAJAN are number of neurons in RBL and methods for locating the centre viz. sample and k-means. In each method, three options explicit, isotropic and k-nearest neighbours are available.

RESULTS AND DISCUSSION

Intelligent Problem Solver (IPS) of Trajan: IPS is an automated procedure in arriving at optimum set of architectures for a data set. It is a fast solution finder even for a novice. The option, 'advanced' is for experts, wherein the types of NNs, number of layers, CPU time limit, performance and training/testing of data sets can be opted based on the need for a task on hand. The results can be exported to Excel, Word or Statistical package for further analysis. The inferences are only a guideline and further runs are to be carried out varying the type of training algorithms, epochs and parameters specific to the network chosen. A summary of the study with IPS and experiments designed now for the data sets with and without outliers and models employing (a) α^2, charge and (b) α and charge follow.

Intelligent problem solver (IPS) of TRAJAN outputted 319 neural network (NN) models employing built in heuristics implemented phase wise to arrive at best possible architectures and performance for the full dataset (NMR01.dat). Its approach is limited however, with a few typical training algorithms and minimum number of epochs. The approximate results are the point of start for a detailed investigation. The modelling and prediction of ^{13}C-NMR response of full data set (NMR01.dat) of halomethanes and that without outliers deleted (NMR04.dat and NMR14.dat) was investigated with a set of feed forward (FF-) multilayer perceptron (MLP), Radial basis function (RBF), Generalized regression (GR) and Linear (Lin) neural networks (NNs). The explanatory variables considered are charge and function of α. Based on correlation (x,y) and percentage explainability of linear model, square of α was used in NMR01.dat and NMR04.dat. Non-linear NNs model (even) complex non-linear profiles and thus in NMR11.dat and NMR14.dat α itself is employed. A perusal of root mean squares (RMS) error and architecture revealed that GR and Lin NNs are of little use and hence a detailed study of RBF and MLP NNs is taken up.

RBF NN model with 22 and 23 neurons for 23 data points resulted in 10^{-13} RMS error. It is strict interpolation RBF model wherein the number of radial layer (RL) neurons is equal to the data points. But, with decrease in number of RL neurons, there is a large increase in RMS error rendering the model inadequate for the current task of function approximation. The modification of transfer function, inclusion of an additional MLP layer, cell growth structure and novelty detection etc are not possible with software employed now. Software modules are underway to develop adoptive RBF-NN to model complex data sets.

SLP-NN with α^2 and charge as input variables

Data set (NMR04.dat) without outliers: The data set NMR04.dat containing variation of ^{13}C NMR response with α^2 and charge as explanatory variables after deletion of statistically detected outliers is analysed by IPS of the software package. The range or band of RMSE of each architecture (2-H1#-1) with a chosen/user select number of neurons throws information on the refinement of NN models phase wise (Figure 2a).

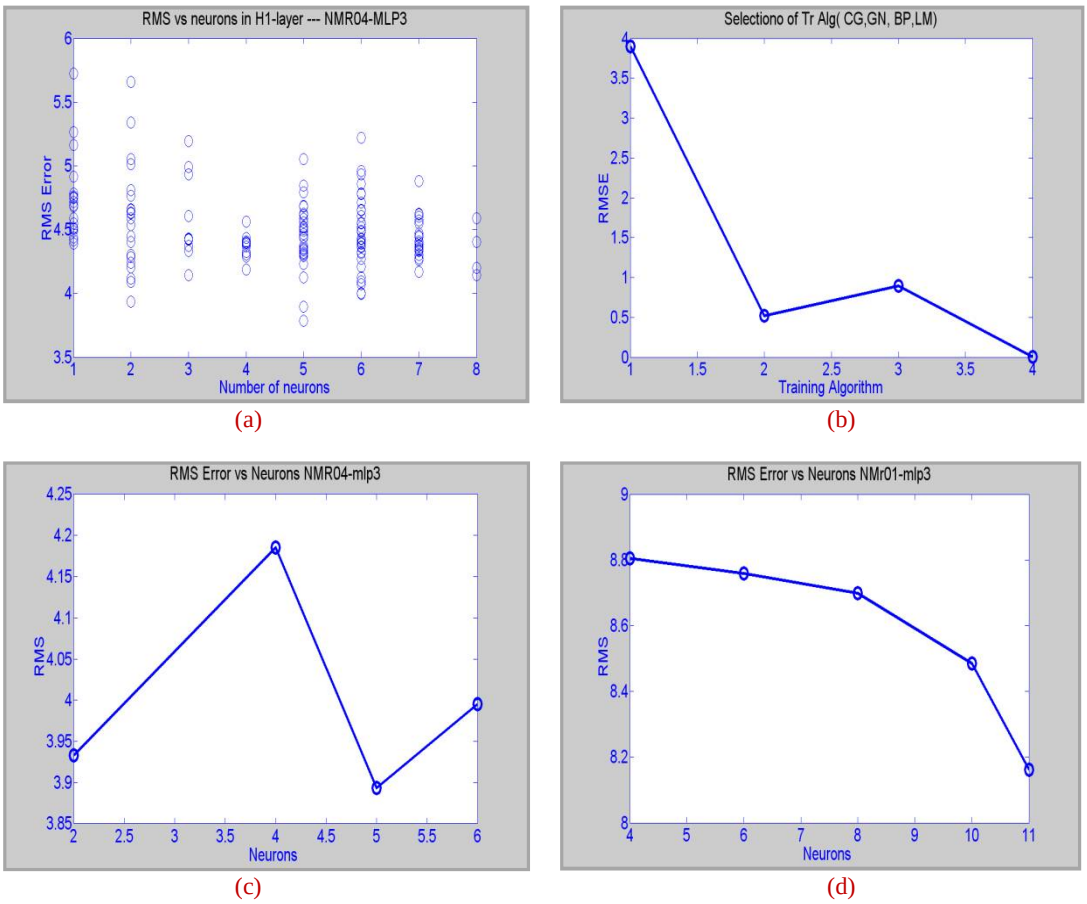


Figure 2. (a) IPS output-effect of H1# neurons on RMS in multiple runs (b) Functioning of training [conjugate gradient (CG), Gauss-Newton (GN), back propagation (BP) and Levenberg Marquardt (LM)] algorithms (c) Effect of neurons on RMSE without data outliers (d) Effect of neurons on RMSE with full data set.

The functioning of optimisation algorithms in arriving at solution (Figure 2b) indicates that Marquardt algorithm (LM) (RMSE: 1.0e-14) is a method of choice over QN (RMSE: 0.52) and BP (RMSE: 0.89). The number of hidden neurons and epochs using LM are studied by one variable at a time (OVAT) procedure. The SLP with 2-5-1 architecture is adequate and that with higher number of neurons is over ambitious indicated by error of order 10^{-14}. The runs to train SLP with LM in duplicate (Figure 3a) and with smaller increments in the near optimum range (Figure 3b) infer that around 12500 epochs bring down RMSE to a value less than 0.01.

Thus, SLP (2-5-1 architecture) with sigmoid transfer function and LM training algorithm (12K epochs) develops an adequate model for the data set without outliers.

Full dataset: The NN computational runs are performed for the full data set (NMR01.dat) including outliers to obtain the optimum NN model (Table 3) by OVAT variation of training algorithm, epochs (Figure 3c, 3d) and number of neurons (Figure 2d) using RMSE as a probe like response of an instrument in a conventional experiment.

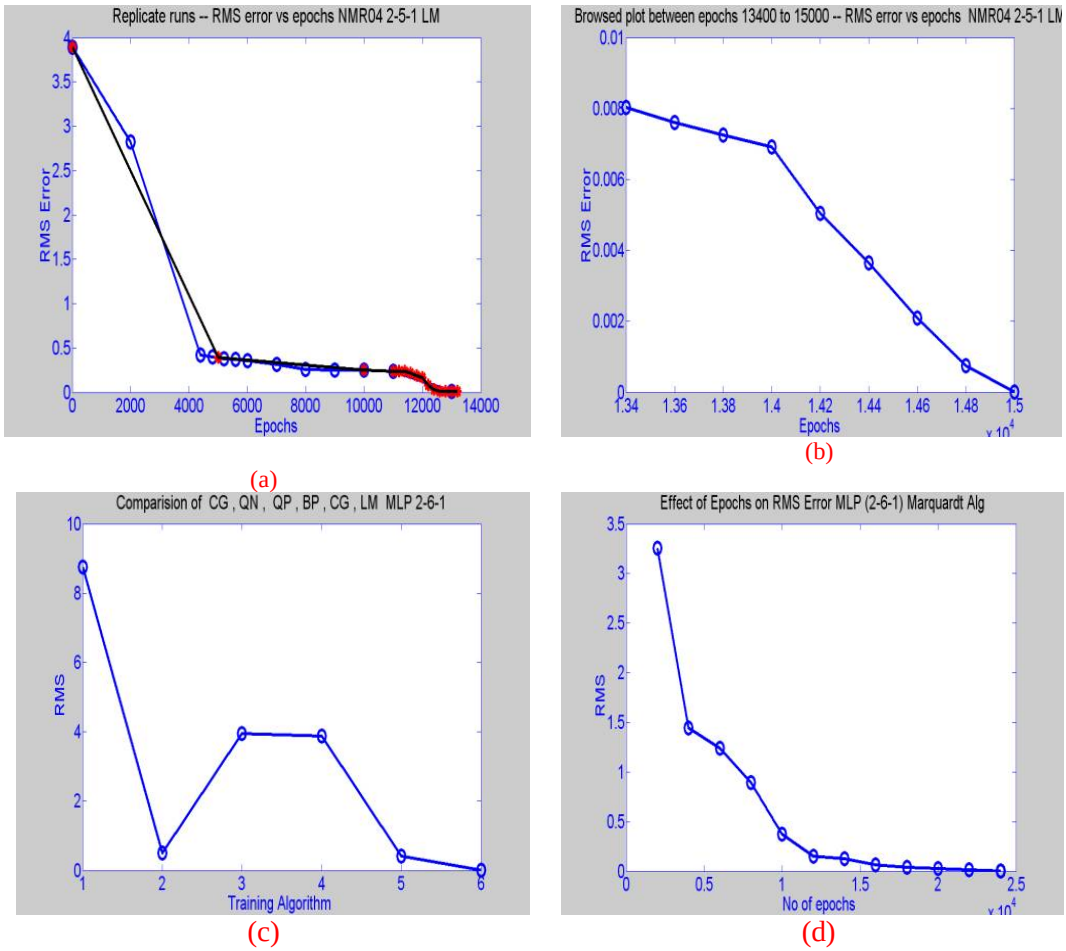


Figure 3. (a) Duplicate runs for the effect of epochs on RMSE with LM (NMR04) (b) Browsed plot of (a) for RMSE vs epochs (NMR04), (c) Effect of Training [conjugate gradient (CG), Quasi Newton (GN,BFGS), Quick propagation (QP), Gauss-Newton (GN), back propagation (BP) and Levenberg Marquardt (LM)] algorithms (d) Effect epochs on RMSE for full data set (NMR01).

Table 3. Optimum models for ^{13}C NMR response data.

(a) NN models

NN	Alpha^2 , Charge				Alpha , Charge			
	NMR04	RMSE	NMR01	RMSE	NMR14	RMSE	NMR11	RMSE
SLP	2-5-1	0.088	2-6-1	0.0027	2-5-1	0.2732	2-9-1	0.26
4-layer MLP	2-5-5-1	0.0998	2-6-2-1	0.0084	2-13-9-1	0.1383	2-13-9-1	0.10
RBF	2-16-1	0.00046	2-19-1	0.0034	2-16-1	0.016	2-21-1	0.017
GRNN	2-18-2-1	0.0346	2-23-2-1	0.01	2-18-2-1	0.57e-3	2-23-2-1	0.0033

(b) Comparison of hard and soft model and data driven procedures

Name		Model	Prediction	Category	
Lin or Lin NN		F	F	Hard	
LMS		F	F	Soft	
PLSR	Lin	F	F	Hybrid (Hard + Soft)	*(a)* ***Model driven***
	Quad	F	F		
	Cubic	F	F		
SLP		T	T		
4-layer MLP		**T**	**T**	Soft	
RBF		T	T		**(i)** **Data Driven**
GRNN		T	T	Theoretical Net	

A comparison of RMSE for full set and that without outliers (Figure 2c, 2d) clearly indicates that more neurons are required for data with outliers. Further, magnitude of entire profile is much lower for pruned data. The results indicate an optimum architecture of 2-6-1 whereby NN could explain the variation of chemical shift of ^{13}C NMR of even statistical outliers with the addition of a few neurons in the hidden layer of SLP. The popular 3-layer MLP also called SLP adequately models NMR01 and NMR04 with α^2 and charge as causative variables.

NNs with non-linear TFs in the hidden layer models even very complex non-linear multi-dimensional surfaces with discontinuities and extrema. The explanatory variables can be directly used in the input layer without employing a functional (log, exp, square root, square, cube, sine, tan etc) transformation. The computational experiments are now designed to propose an NN model for the linear or first order form of α and charge.

SLP-NN with α and charge as input variables

Data set (NMR14.dat) without outliers: The adequate SLP architecture found when α and charge are the input variables is 2-5-1 (Figure 4a, 4d). A minimum of around 40K epochs are required with LM optimisation procedure (Figure 4b, 4c). The other training algorithms Gauss Newton (GN), quasi Newton (QN) alone or in combination could not surmount local minima at higher RMSE. These results show that 2-5-1 (Lin-logistic-Lin) SLP and LM optimisation algorithm successfully explains the variation of chemical shift of ^{13}C NMR with α and charge also.

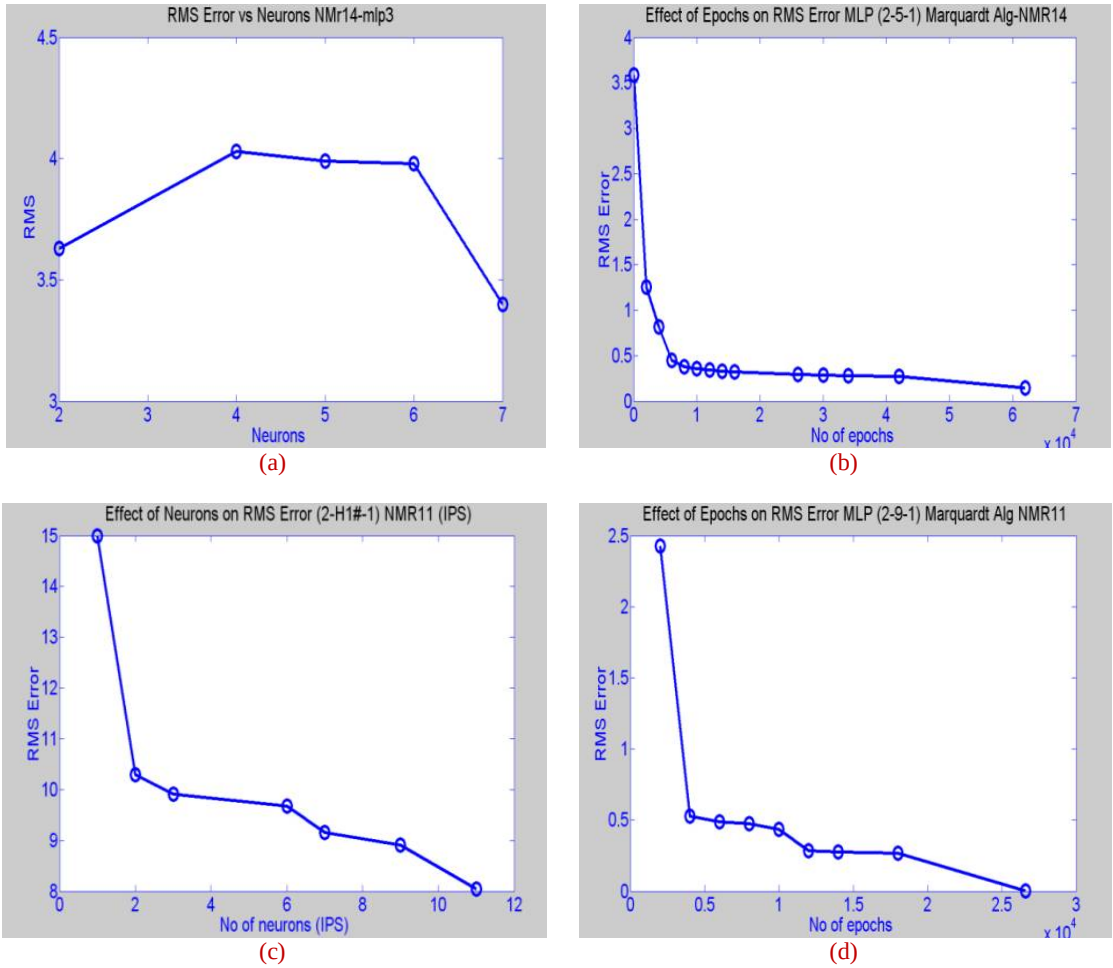


Figure 4. (a) Effect of neurons on RMSE with LM (NMR14) (b) RMSE vs epochs (NMR14) with LM Training algorithm (c) Optimum number of H1# neuron (NMR11) from IPS (d) RMSE vs epochs (NMR11) with LM Training algorithm for full data set

Full data set (NMR11.dat with outliers): A detailed study of the number of hidden layer neurons and epochs for training the full data set with variables α and C resulted in 2-9-1 SLP as an adequate NN model for the response. The results of SLP with four data sets planned unequivocally established that it models NMR response with tolerable residuals with a range of 5 to 9 hidden neurons depending upon that data structure and functional form of input variables. The addition of one or more hidden layer to SLP results in multi (4- or m-) layer perceptron (MLP) NNs, which account for complex non-linearities of response profile. With this perspective experiments with 4-layer sigmoid MLP were performed.

Four-Layer MLP-NNs: The TRAJAN runs are performed for all the four data sets (discussed in SLP section) with 4-layer MLP-NN employing sigmoid TF and varying the parameters of architecture and training algorithms. The variation of neurons/PEs in the two hidden layers (H1#, H2#) of MLP-NN is studied with the a priori information from SLP results. The data set NMR14.dat with α and Charge as variables after elimination of outlying data points, MLP 2-6-7-1 (RMSE: 0.0457) is adequate while 2-2-1-1 (RMSE: 2.156) and 2-13-9-1 (RMSE: 0.1383) (Figure 5a) are inadequate. With outliers, the optimum architecture is 2-13-9-1. In this case, both QN as well as LM trained well with an RMSE < 0.1 (Figure 5b, 5c). However, BP (Figure 5d), QP and DBD and CG could not find the optimum even after 500K epochs and even by jogging the weights.

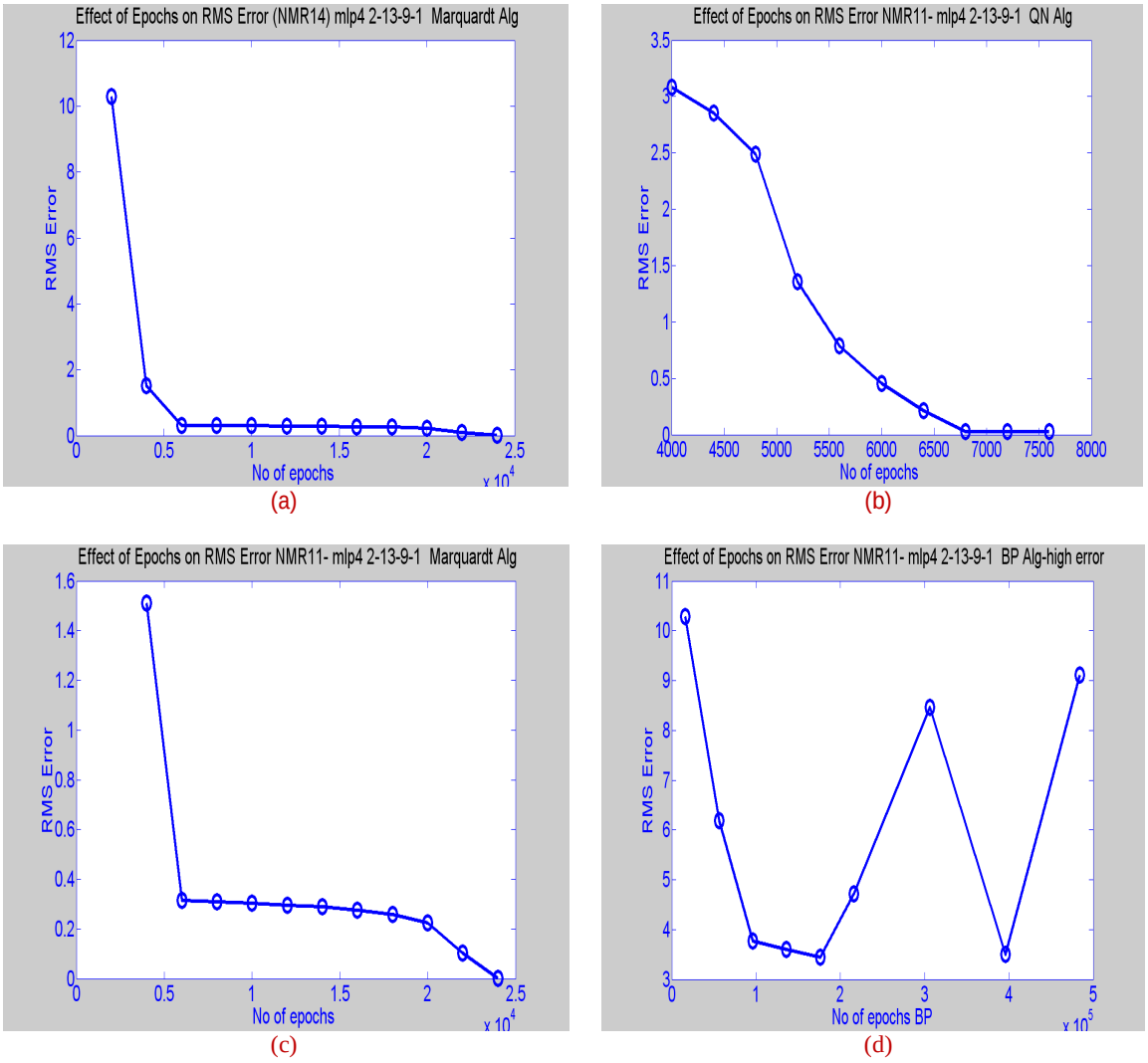


Figure 5. (a) Effect of epochs on RMSE with LM (NMR14) (b) RMSE vs epochs (NMR11) with QN training algorithm (c) RMSE vs epochs (NMR11) with LM (d) RMSE vs epochs (NMR11) with BP

MLP 2-5-5-1 and 2-6-2-1 models are found to be acceptable as expected when α^2 is used instead of α. This is obvious from the fact that α^2 itself imbibes non-linearity and thus smaller number of hidden neurons is able to model. The effect of epochs in training 2-6-2-1 MLP with QN and LM revealed that initially QN has lower RMSE, but both converge after 10K of epoch.

RBF NNs: An inspection of the bar diagram (Figure 6a) shows the range of RMSE for each RBF architecture (#RL neurons) with outliers and function of α. It is inferred that α^2, Charge models have lower RMSE compared to those of α and Charge. Further, the data set without outliers has lower RMSE than those with full data sets. The RL neurons in the models with adequate explainability are nearly equal to number of patters in the data set, rendering the procedure to be strict interpolation RBF (SI_RBF). The selection of methods to locate centre is studied by monitoring RMSE for different options by OVAT procedure for 2-16-1 RBF-NN with fixed deviation of 1.0 and restricting the number of neighbours to three for the data set NMR11.dat. Sample-explicit, sample-isotropic and k-means-explicit are comparable. The optimum number of nearest neighbours and width of Gaussian kernel are 3 and 1.0 (Figure 6b, 6c).

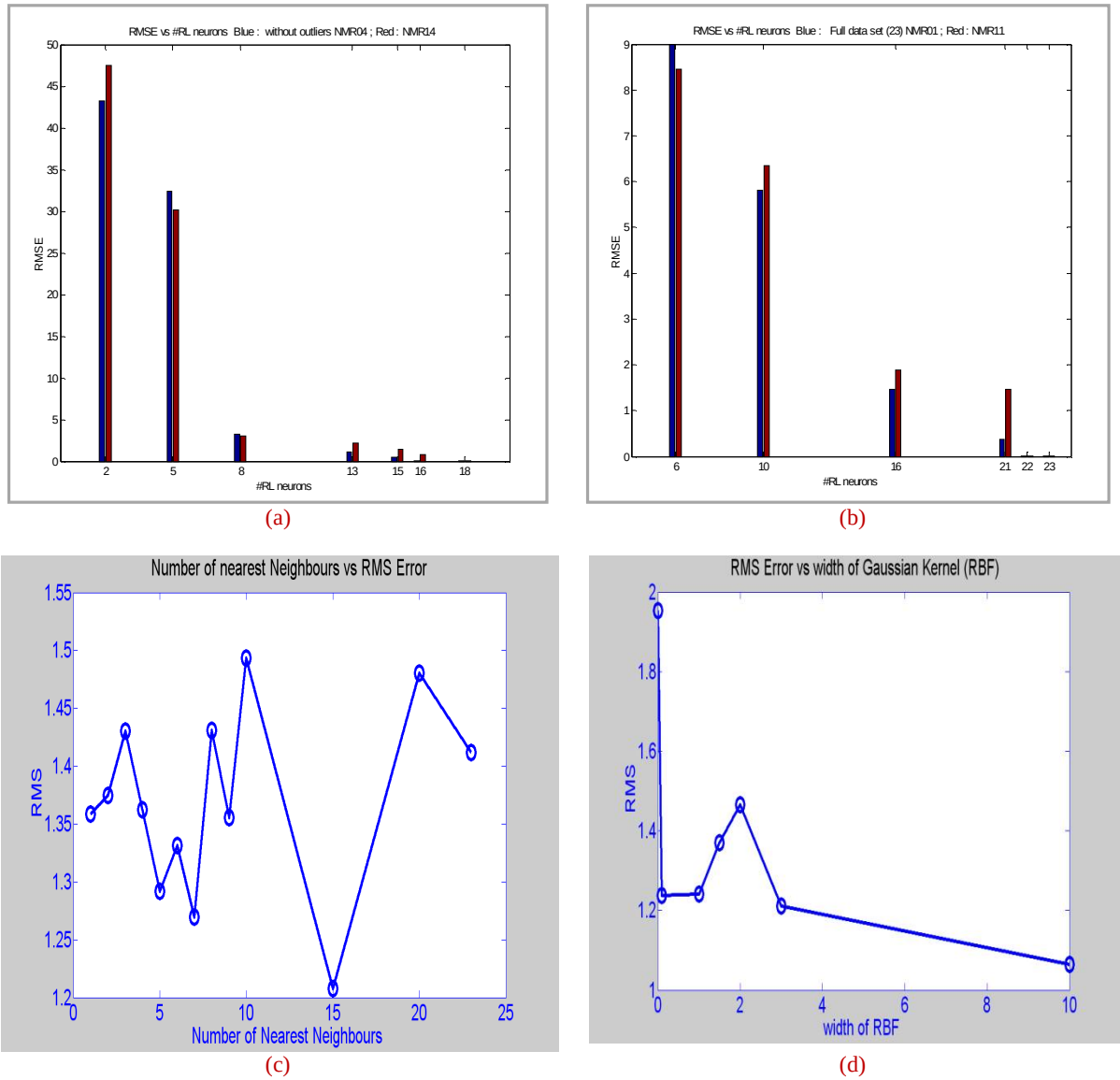


Figure 6. Comparison of RMSEs RBF-NNs with RL neurons, (a) Without outliers Blue : alpha^2 ; Red : alpha, (b) Full data set Blue : alpha^2 ; Red : alpha, (c) Effect of nearest neighbors on RMSE in training RBF-NN, (d) Effect of Kernel width on RMSE in RBF-NN.

GRNN: The variation of smoothing factor (SF) vs RMSE in GRNN (Figure 7) shows that the range of SF 0.01 to 0.02 results in tolerable ^{13}C NMR response. For data sets without outliers (NMR04.dat and NMR14.dat) the trend of RMSE is similar for both (i) α^2 vs charge and (ii) α vs charge models. With full data set obviously the RMSE curves are a little far away from others. Even though GRNN is based on the theory of probability and Bayesian statistics, it could model ^{13}C NMR response within acceptable range of residuals unlike unweighted MLR or LMS.

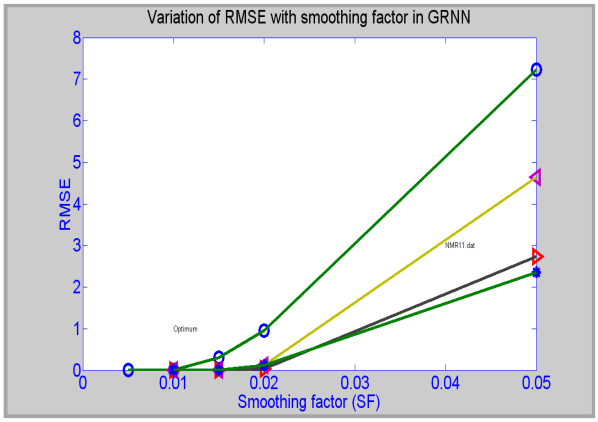


Figure 7. Effect of smoothing factor of GRNN on RMSE.

Lin (ear)-NN: The residuals in ^{13}C NMR chemical shift with one and two variables for the four data sets (NMR01.dat, NMR04.dat, NMR11.dat and NMR14.dat) show that both the variables explain the variance, yet the model is inadequate and invalid. The architectures used in this study are depicted in figure 8.

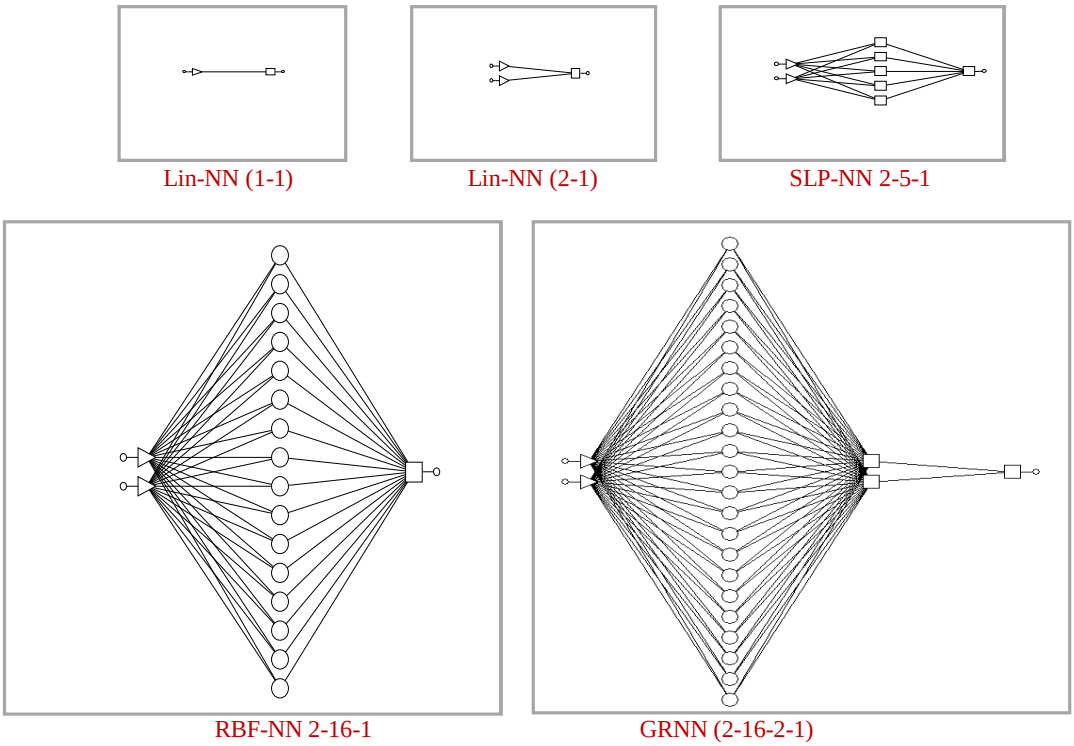


Figure 8. Architectures of different NNs for analysis ^{13}C NMR response.

Test data sets: Exhaustive test data sets were developed to study the predictive (interpolation, and extrapolation) capability. The data sets deleting the test patterns are trained with the optimum

architectures obtained with all the points in that data set. The RMSE of test sets are obviously higher than that for training (Table 4). The residuals for different test data sets are within tolerable limits, although the magnitude is higher than that for training dataset and with experimental reproducibility.

Table 4. Prediction of ^{13}C NMR response with Neural Network model

Explanatory variables	Neural Net		File Name	RMSE in response		Point Number	Response			NN number	
	Name	Arch		Training	Test		NNcal	Exp	Res		
Alpha	charge	MLP	2-5-1	nmr14	0.006796	0.3411	2	71.91	71.57	**0.34**	212
Alpha	charge	MLP	2-5-1	nmr14	0.3483	1.095	5	44.49	43.4	**1.09**	215
Alpha	charge	MLP	2-5-1	nmr14	0.001998	0.70885	7	25.59	26.3	**-0.70**	217
Alpha	charge	MLP	2-5-1	nmr14	0.03085	0.9714	9	81.32	82.3	**-0.97**	220
Alpha	charge	MLP	2-5-1	nmr14	0.7722	0.5594	15	72.34	72.9	**-0.56**	223
Alpha	charge	MLP	2-5-1	nmr14	0.2073	0.8509	16	40.85	40.0	**0.85**	225
Alpha	charge	MLP	2-5-1	nmr14	0.6645	0.4937	17	9.50	10.0	**0.5**	226

Explanatory variables	Neural Net		File Name	RMSE in response			NP	Response			NP	Response			
	Name	Arch		Train		Test		NN	Exp	Res		NN		Exp	Res
Alpha ^2	charge	MLP	2-5-1	nmr01	0.1068	0.2934	7	58.3	58.6	0.3	12	82.3	82.58	0.28	194
Alpha ^2	charge	MLP	2-6-2-1	nmr01	0.228	0.3695	7	58.42	58.3	0.125	12	81.79	82.3	-0.5	184

The results for 4-layer and 3-layer MLP-NNs even with (i) α^2 vs charge and (ii) α vs charge as variables have the same trend, establishing the high predictive ability of NNs for NMR chemical shifts.

Analysis of 13C NMR spectral tasks in the mathematical framework: Identification of compounds from limited number of ^{13}C NMR chemical shifts from a smallest set of compounds is the classical look up table method. When, large (computer readable) database is used, similarity measures and knowledge base are employed. On the other hand, full spectrum comparison is a pattern recognition task either in supervised or unsupervised mode. The classification, discrimination and clustering algorithms find place for a series of homologous compound sets of substituents, fragments and/or moieties.

The variations of ^{13}C NMR chemical shifts of a series of compounds are modelled with macroscopic and microscopic physico-chemical parameters and latest molecular (quantum-chemical/topological/electrochemical) descriptors with apriori knowledge of cause-and-effect relationship. Here, the linear and non-linear (hard and soft) regression methods in explanatory (chemical) /orthogonal (mathematical transformed) space are in vogue in yester years.

The cause-and-effect model is solved with two prime objectives viz. minimum residuals and statistically significant model parameters. In curve fitting, even outliers and /or those with high (measurement) error are also fitted at the expense of the validity of model parameters. On the other hand, in parameterisation task physico-chemical/statistical significance of the parameters is retained although the residuals are high for (i) singleton clusters and (ii) patterns with different trend. Linear, quadratic, cubic and polynomial of higher order (<8) or Poisson, log normal regression models based on the shape of the scatter profile are the model driven techniques well nurtured over a period of 50 years.

Complex multi-dimensional surfaces with discontinuities, multiple extrema are common in real large/small data sets and data driven information technologies like NNs are sought-after present-day paradigms. NNs with latest improvements never fail [**30, 31**] in spite of the fact the performance is degraded with complexity of the profile, paucity of data, high noise and outliers.

Thus, the variation of ^{13}C NMR response with α and charge is a function approximation problem. This task is solved in NN paradigm with RBF and MLP and GRNN philosophies.

State-of-the art of modelling: The linear correlation coefficient in explanatory variable space is crucial to prune X-variables and indicate the need for orthogonalisation to alleviate the problems of multi collinearity [27c]. The correlation between response space and explanatory variable space, ANOVA and multiple-ANOVA (MANOVA) reflect explainability of the variable of response with the model. The validity of regression coefficients endorses the acceptability of the models for generalisation with a possibility to interpret the regression parameters on a physico chemical basis. The residual analysis rules out the outliers, trend (an artefact of model error) and the dependability of the model within the measurement precision. The cross validation (leave one out (LOO), leave k-out (L-kO)) and influential statistics estimates the efficacy of data points on the model. The gross limitations of these hard parametric models are that no real data set adheres to the necessary conditions of distributions of regression parameters and the presence of normal noise only in the response. They however paved way to non-parametric and distribution free methods and finally to data driven evolutionary paradigms. The current trend is to employ hybrid technologies in hierarchical and sequential manner with due emphasis on earlier well-established theoretical models coupled with unexplainable but indispensable black box/ empirical best set of models. In spite of the unsolved riddles of both the worlds, the hybrid models have been unequivocally found superior in providing complementary and supplementary information to alter/ design new experiments. The small number of data points and tradition driven low dimensional space on one hand, terabytes of data in 100 to 1000 dimensions on the a other hand, plethora of computational tools, software packages, accumulated knowledge/traditional commonsense bases etc, added a dilemma to the scientist rather than arriving at even confident set of solutions. But, it leads to new paradigm implementable with ease on future hard ware. A statistically valid model generated I/O mapping and is acceptable for interpolation and with a caution for inverse interpolation. The prediction (forecast and hind cast) is inevitable and the errors depend upon several factors, the prime being validity of the model in the rage of prediction.

^{13}C NMR Dataset: The residuals for the typical data set using MLR and PLSR (Supl-Figure 1) amply show that these hard and soft model driven procedures are inadequate. A typical screen dump of TRAJAN with 3-layer MLP for the analysis is depicted in figure 9.

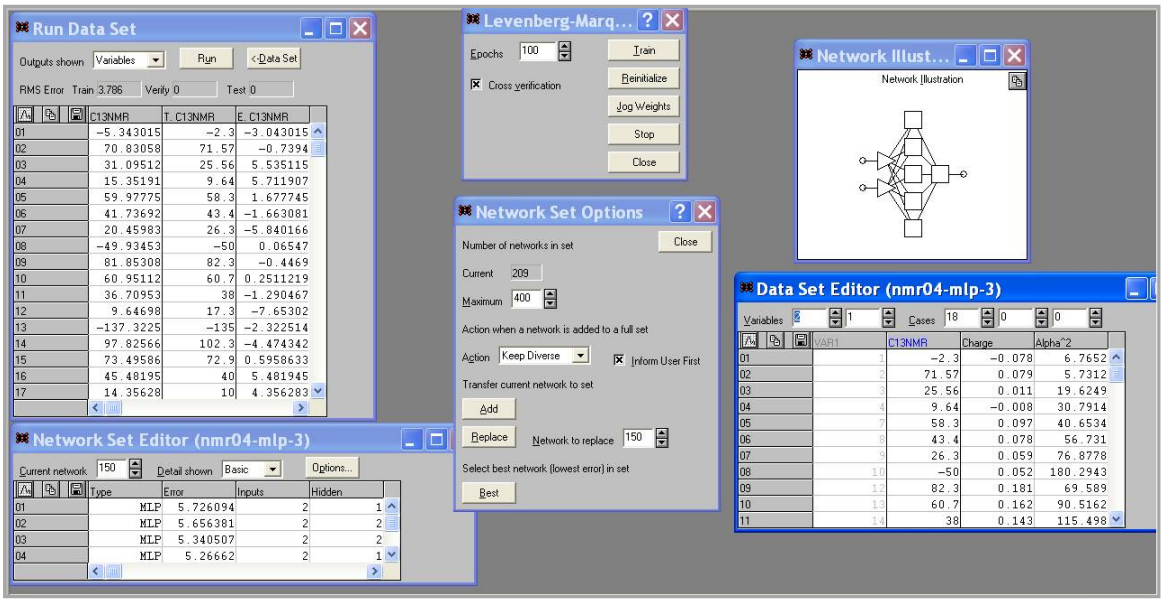


Figure 9. Typical screen dump of Trajan (MLP 2-5-1 architecture) for analysis of NMR04 data set.

A rigorous study of residuals of NN models was made and (Figure 10a, 10c) incorporating the residuals for SLP 2-5-1 shows the efficacy of NN paradigm. As the maximum value of residual is less than 0.6 for data set with and without outliers (detected by LMS), NN is versatile tool to model ^{13}C response in the hands of organic chemists. The very high magnitude (30 to 100) of residuals with linear-NN (Figure 10b, 10d) or multiple linear least squares suggest NN is indispensable both for modelling and prediction. Increase in number of epochs, hidden layers/neurons although decrease RMSE, optimum parameters are chosen such that noise is not fitted in the model.

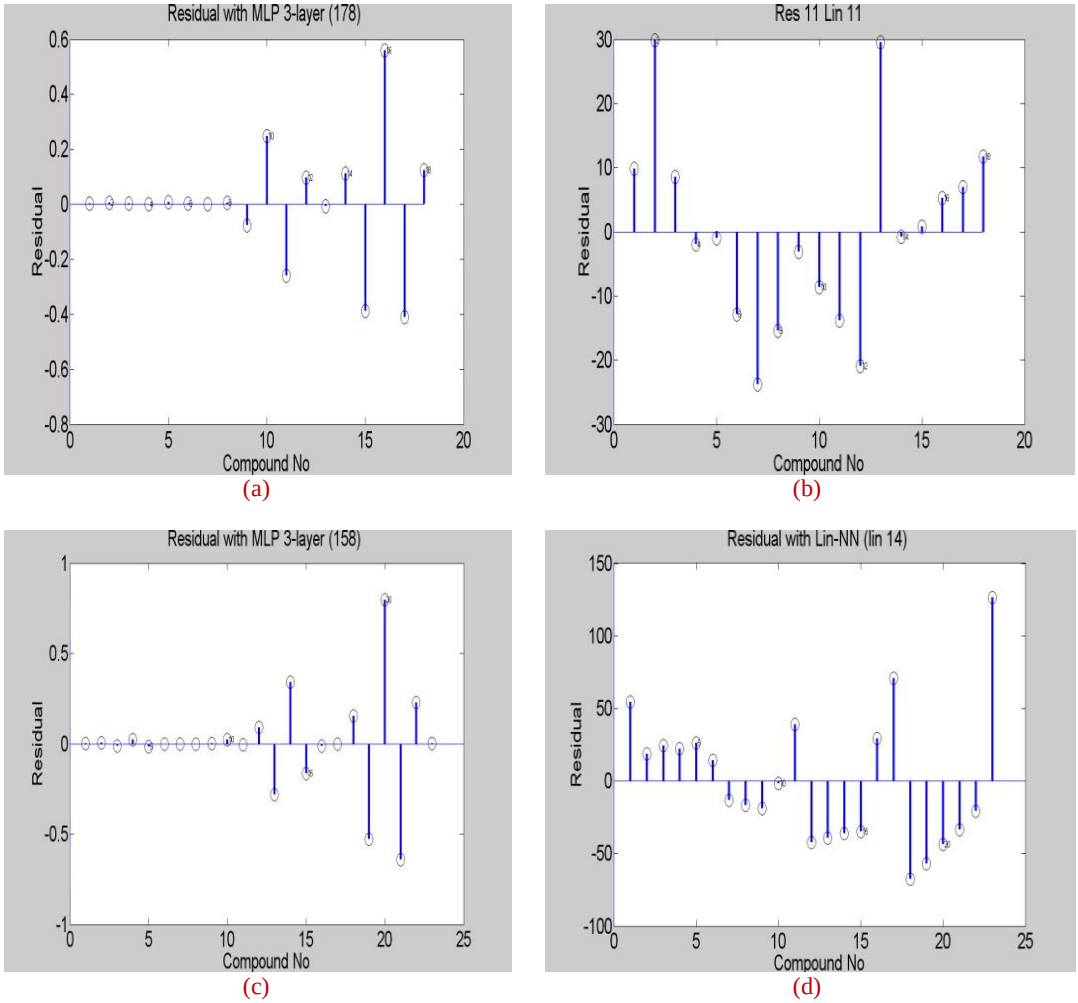


Figure 10. Residual plots (a) MLP-3 layer-NN without outliers (b) Lin-NN without outliers (c) MLP-3 layer-NN full data set (d) Lin-NN full data set.

The initial weights in training algorithms of NNs are based on random numbers and thus the end numerical results slightly differ. Replicate runs for the variation of epochs show the trend of convergence in the training for a chosen architecture is the same. Throughout the investigation at least duplicate runs are performed for training algorithms and IPS. The range of RMSE with neurons in the output of IPS is a cumulative effect of multiple training algorithms used in sequence, number of epochs and initial weights. However, a study of the range of chosen architecture is more informative about the precision of RMSE.

A perusal of the results (Table 3) establishes that a variety of NN architectures (Figure 8) based on different philosophies RBF, MLP and GRNN could model and predict ^{13}C NMR chemical shifts of 23 halomethanes. The primary goal is to make a feasibility study for the use of NN tool in NMR spectroscopy and not to attempt a critical comparison of the features including limitations of different

NN models. This is limited by the size of the experimental data set and non-availability of recent modifications in NN at various phases in the software form.

Three-layer MLP i.e. SLP is preferable to 4-layer MLP based on parsimony principle. RBF model is here in near strict interpolation mode. Of course, the number of neurons in radial layer can be decreased either by using raised cosine TF, growing cell structure [**59**], higher order neurons or novelty detection modules [**60**]. The fact that GRNN is an adequate model can be made use with larger data set in reaping the benefits of using priori/posterior probabilities in the sub goals of finding the centres of clusters. This does not in any way hamper the NN basis and the results are still independent of limitations of classical statistics. The model driven parameterised generalisation starts with simple as possible (SAP) and polynomial/exponential/ transcendental terms are added based on the trend in residuals of consecutive data points with respect to response or explanatory variables. But in data analysis with an a priori (multi-linear) model, the data with high residuals in response are eliminated based on statistical outlier detection procedures. Here, the ranges of scaled and MAD statistics are employed. In the case of experimental data of high precision and accuracy statistical validation is a coarse filter and requires chemical fine-tuning. Thus, a statistically validated adequate model is subjected to peer scrutiny within the framework of measurement reproducibility. The elimination/setting apart of data subsets are indispensable for many chemical tasks. The analysis with multiple responses (^{13}C, ^{1}H), molecular descriptors (CODESSA) with larger number of halo substituted hydrocarbons using fuzzy ARTMAP and modular-NN is in progress.

ACKNOWLEDGEMENTS

The author (RSR) thanks Department of Ocean Development for financial support for the project on predictive models to procure the software TRAJAN and Profession II of neuralware.

Supplementary Information:

Sup-Table 01. LMS and LS Regression parameters and sdy for data sets with elimination of outliers

(a) LMS procedure

	NMR01	NMR02	NMR03	NMR04	NMR11	NMR12	NMR13	NMR14
a0	37.606	39.487	39.487	*39.487*	95.519	106.23	91.131	*91.131*
a1	476.77	480.66	480.66	*480.66*	674.53	648.02	668.2	*668.2*
a2	-0.64576	-0.63497	-0.63497	*-0.63497*	-16.099	-16.473	-15.883	*-15.883*
sdy	2172.9	84.304	41.927	*22.466*	7897.2	939.53	404.23	*409.34*

(b) MLR procedure

	NMR01	NMR02	NMR03	*NMR04*	NMR11	NMR12	NMR13	*NMR14*
a0	51.597	38.615	36.506	*38.879*	101.79	117.36	88.618	*88.381*
a1	245.87	509.03	496.5	*482.6*	225.19	666.52	577.94	*580.37*
a2	-0.57445	-0.67306	-0.63688	*-0.63583*	-12.29	-18.127	-13.68	*-13.698*
sdy	28.02	6.9663	6.1033	*4.7083*	48.226	29.131	15.485	*15.982*

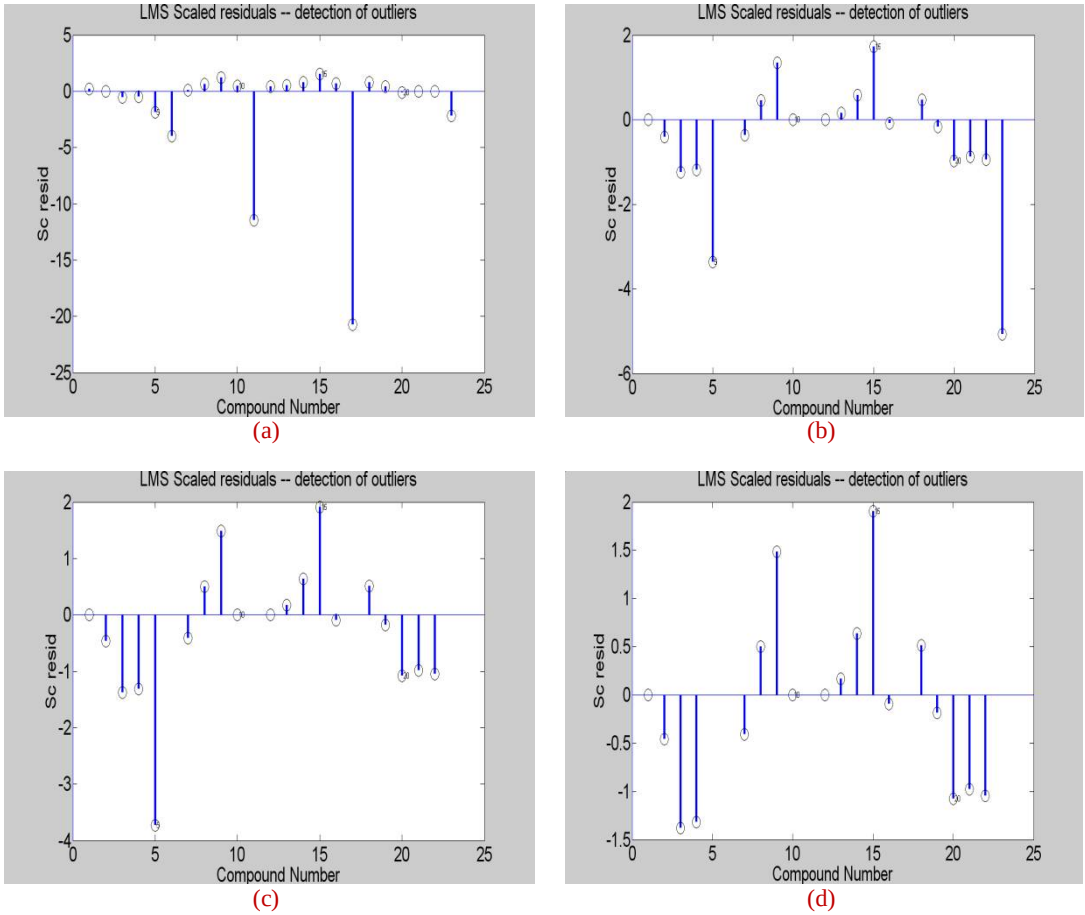


Figure 1 Sup. Detection of outliers with LMS scaled residual for alpha^2 and charge model (a) NMR01 (b) NMR02 (c) NMR03 (d) NMR04.

Sup-Table 2. Detection of outliers with LMS Scaled residuals

Compound Number	II. LMS SCALED RESIDUALS			
	NMR01	NMR02	NMR03	NMR04
1	0.19293	-4.9029e-016	-5.4398e-016	-5.4336e-016
2	0	-0.41409	-0.45943	-0.4589
3	-0.53959	-1.2427	-1.3787	-1.3771
4	-0.49874	-1.1869	-1.3169	-1.3154
5	-1.8423	-3.362	*-3.7301 (*)*	NaN
6	*-4.0077 (*)*	NaN	NaN	NaN
7	0.081737	-0.36759	-0.40784	-0.40738
8	0.61239	0.44965	0.49889	0.49832
9	1.193	1.3375	1.484	1.4823
10	0.47079	0	0	0
11	*-11.481 (*)*	NaN	NaN	NaN
12	0.38984	2.6149e-015	2.9012e-015	2.8979e-015
13	0.5035	0.15102	0.16756	0.16737
14	0.79459	0.57337	0.63615	0.63542
15	1.5492	1.7174	1.9055	1.9033
16	0.67127	-0.086132	-0.095563	-0.095454
17	*-20.742 (*)*	NaN	NaN	NaN
18	0.76904	0.46318	0.5139	0.51331
19	0.39325	-0.1654	-0.18351	-0.1833
20	-0.088648	-0.96905	-1.0752	-1.0739
21	1.6607e-015	-0.88257	-0.97921	-0.97809
22	4.1516e-016	-0.94384	-1.0472	-1.046
23	-2.1681	*-5.0806(*)*	NaN	NaN

Sup-Table 3(a). PLS regression characteristics with for data sets with elimination of outliers: Standard deviation in ^{13}C NMR response (y)

Parameter	SD in ^{13}C NMR response							
	Alpha^2 & Charge				Alpha & Charge			
	NMR01	NMR02	NMR03	NMR04	NMR11	NMR12	NMR13	NMR14
PLSC1	80.3756	72.7651	59.214	60.6836	90.32	84.5837	58.4945	59.6961
PLSC2	41.5214	24.9478	22.7087	23.0846	66.41	62.438	40.5776	41.1752
PLSC1 ^2	28.1809	19.9639	18.8047	19.3527	20.6105	10.1531	10.1465	8.4785
PLSC2 ^2	20.9884	19.787	17.625	17.7509	11.0112	10.3709	10.3972	8.1928
PLSC1*PLSC2	20.8888	18.2271	16.9298	17.0982	11.2485	10.4689	10.5617	8.2384
PLSC1^3	17.4262	16.1511	15.3742	14.3864	10.1283	9.7324	8.9042	6.6208
PLSC2^3	17.2094	14.2446	15.7224	14.9073	10.3703	10.0443	9.2677	6.8199

Sup-Table 03(b). PLS regression characteristics with for data sets with elimination of outliers
(a) Regression coefficients and their standard deviations

Parameter (SD par)	Alpha2 & Charge				Alpha and Charge			
	NMR01	NMR02	NMR03	NMR04	NMR11	NMR12	NMR13	NMR14
PLSC1	-0.40055 (2.425)	-0.2203 (1.8855)	-0.65844 (3.5669)	-0.8226 (3.523)	-15.9498 (17.4756)	-14.8917 (16.6746)	18.5379 (20.6822)	18.3225 (11.2107)
PLSC2	-485.6943 (1822.7232)	-558.047 (1393.6722)	-480.8471 (1850.6143)	-401.6424 (1817.65)	-437.2552 (653.1424)	-533.5439 (1167.3126)	-484.6773 (963.802)	-487.1556 (510.5652)
PLSC1 ^2	-0.0030714 (0.013972)	-0.0021954 (0.010544)	-0.0052698 (0.034817)	-0.0066386 (0.033783)	-1.3507 (2.2261)	-1.299 (2.1085)	-1.869 (3.3294)	-1.7511 (1.813)
PLSC2 ^2	23.3069 (9760.4238)	2811.7072 (16248.0381)	597.4535 (14698.5815)	485.1032 (12897.2097)	-421.0796 (3046.4458)	241.0849 (5656.2881)	-62.1109 (3995.4251)	287.7779 (2153.9954)
PLSC1* PLSC2	-0.58307 (10.9529)	-1.2501 (8.9888)	0.11826 (15.4657)	0.32764 (14.1305)	2.2722 (71.9882)	-4.823 (90.3906)	-0.70419 (78.5973)	0.27069 (42.5324)
PLSC1^3	-2.7712e-006 (1.6305e-005)	-9.8892e-007 (1.4391e-005)	-7.7207e-006 (8.8385e-005)	-1.0565e-005 (8.3999e-005)	-0.013599 (0.068423)	-0.011325 (0.064966)	0.035744 (0.1337)	0.02897 (0.073655)
PLSC2^3	935.0923 (13452.0858)	12989.7169 (82764.1865)	4722.2348 (113139.7178)	2874.9197 (102170.0827)	216.7449 (4838.2262)	2277.0228 (60263.3638)	-112.7497 (54411.5879)	2500.1391 (30655.4918)

REFERENCES

[1]. H. Yoshida, Y. Miyashita, S. Sasaki, *Chemom Intell Lab Syst.*, **1996**, l32 , 193.

[2]. B. Wendt, U. Uhrig, F. Bos, *J Chem Inf Model*, **2011**, 51, 843.

[3]. J. Hao, X. Zou, M. Wilsonc, N. P. Daviesc, Y. Sunc, A. C. Peetc, T. N. Arvanitisa, *NMR Biomed.*, **2012**, 25, 594–606.

[4]. E. Mosera, F. Stahlbergd, M. E. Ladde, S. Trattnig, *NMR Biomed*, **2012**, 25, 695. (b) K. Ugurbil, Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project, *NeuroImage*, **2018**, 80, 80-104. (c) B. Ittermann, A 20-T superconducting magnet for magnetic resonance imaging of human brain at ultrahigh field. Magnetic Resonance in Medicine, **2016**, 76(3), 943-952. (d) Budinger, F. Thomas; D. Bird, Mark; Frydman, Lucio; Long, R. Joanna; Mareci, H. Thomas, Rooney, D. William; Rosen, Bruce; Schenck, F. John; Schepkin, D. Victor, A. Sherry, Dean; Sodickson, K. Daniel, Springer, S. Charles, Thulborn, R. Keith, Uğurbil, Kamil, Wald, L. Lawrence, **2016**. Toward 20 T magnetic resonance for human brain studies: opportunities for discovery and neuroscience rationale. Magnetic Resonance Materials in Physics, Biology and Medicine, doi:10.1007/s10334-016-0561-4

[5]. D. Yua, K. Yuana, L. Zhao, L. Zhao, M. Dong, P. Liu, G. Wanga, J. Liu, J. Sun, G. Zhou, K. M. von Deneen, F. Liang, W. Q. J. Tiana, *NMR Biomed*, **2012**, 2, 806.
[6]. H. Witjes, A. W. Simonetti, L. Buydens, *Anal Chem.*, **2001**, 73, 551A.
[7]. M. K. Paira, T. K. Mondal, D. Ojha, A. M. Z. Slawin, E. R. T. Tiekink, A. Samanta, C. Sinha, *Inorg Chim Acta*, **2011**, 370, 175.
[8]. J. S. Anderson, D. M. LeMaster, *Biophys Chem*, **2012**, 168-169, 28.
[9]. V. V. Bardin, *J. Fluorine Chem*, **1998**, 89, 195.
[10]. J. R. Nanney R. E. Jetton, C. A. L. Mahaffy, *J Fluorine Chem*, **1996**, 80, 105.
[11]. D. D. Ska, E. M. Serwick, A. Drelinkiewicz, Z. Rutkowska, D. bika, M. Witko, R. Socha, M. Zimowska, Z. Olejniczak, *Appl Catal A.*, **2012**, 427– 428, 16.
[12]. V. Conte, F. Di Furia, S. Moro, *J Mol Catal A: Chem.*, **1995**, 104, 159.
[13]. T. Geelen, L. E. M. Paulis, B. F. Coolen, K. Nicolay, G. J. Strijkers, *NMR Biomed*, **2012**, 25, 953.
[14]. B. F. Coolen, L. E. M. Paulis, T. Geelen, K. Nicolay, G. J. Strijkers, *NMR Biomed*, **2012**, 25, 969.
[15]. D. L. Flumignan, R. Sequinel, R. R. Hatanaka, N. Boralle, J. E. de Oliveira, *Fuel.*, **2012**, 99, 180.
[16]. P. E. Thelwall, N. E. Simpson, Z. N. Rabbani, M. D. Clark, R. Pourdeyhimi, J. M. Macdonald, S. J. Blackband, M. P. Gamcsik, *NMR Biomed*, **2012**, 25, 271.
[17]. I. S. Flores, M. S. Godinho, A. E. de Oliveira, G. B. Alcantara, M. R. Monteiro, S. M. C Menezes, L. M. Liao, *Fuel*, **2012**, 99, 40.
[18]. P. Mazzei, A. Piccolo, *Food Chem*, **2012**, 132, 1620.
[19]. E. Fereydunia, M. K. Rofouei, M. Kamaee, S. Ramalingamd, S. M. Sharifkhani, *Spectrochim Acta, Part A* **2012**, 90, 193.
[20]. K. A. Blinov, Y. D. Smurnyy, T. S. Churanova, M. E. Elyashberg, A. J. Williams, *Chemom Intell Lab Syst*, **2009**, 97, 91.
[21]. M. U. Roslund, E. Säwén, J. Landström, J. Rönnols, K. H. M. Jonsson, M. Lundborg, M. V. Svensson, G. Widmalm, *Carbohydr Res*, **2011**, 346, 1311.
[22]. K. Endo, T. Ida, S. Shimada, J .V. Ortiz, K. Deguchi, Shimizu T & Yamada K, *J Mol Struct.*, **2012**, 1027, 20.
[23]. C. D. Agostino, M. D. Mantle, L.F. Gladden G. D. Moggridge, *Chem Eng Sci.*, **2012**, 74, 105.
[24]. Y. Xu, W. Qian, Q. Gao, H. Li, *Chem Eng Sci.*, **2012**, 74, 211.
[25]. J. Meiler, W. Maier, M. Will, R. Meusinger, *J Magn Reson*, **2002**, 187, 243.
[26]. A. Braibanti, R. Sambasiva Rao, V. Ananta Ramam, G. Nageswara Rao, V. V. Panakala Rao, *Ann. Chim. (Rome*), **2005**, 95, 291.
[27]. K. Sambasiva Rao, R. Sambasiva Rao, *J. Korean Phys Soc.*, **1998**, 32, S1850. (b) K. Viswanath, R. Sambasiva Rao, Ch. V. Kameswara Rao, K. Rama Krishna, B. Rama Krishna, G. E. G. Santhosh, *J. Applicable Chem*, **2012**, 1, 109. (c) R. Sambasiva Rao, G. Nageswara Rao, Computer applications in Chemistry, Himalya Publisher, New Delhi (India) **2005**.
[28]. J. Gasteiger, I. Suryanarayana, *J. Magn Reson Chem*, **1985**, 23, 157.
[29]. T. Clark, R. Rauhut, A. Breindl, *J. Mol. Modell*, **1995**, 1, 22.
[30]. C. T. Leondes, Neural Network systems, techniques and applications, Vol 1 - 7, edn, Academic press, New York, **1998**.
[31]. P. V. Yee, S. Haykin, Regularized radial basis function networks, theory and applications, John Wiley sons, New York, **2001**.
[32]. M. Jalali-Heravi, S. Masoum, P. Shahbazikhah, *J Magn Reson*, **2006**, 171, 176.
[33]. The Math works Inc., 24 prime park way, Natik, Massachusetts, 01760, USA.
[34]. Trajan Software Ltd, Trajan House, 68 Lesbury close, Chester-le street co., Durham, DH2, 3SR, England.
[35]. A. Braibanti, G. Nageswara Rao, S. B. Jonnalagadda, D. Sudarsan, R. Sambasiva Rao, *Ann. di. Chim.* (Rome), **2001**, 91, 29.

[36]. H. A. R. Laila, K. Sambasiva Rao, G. Nageswara Rao, R, Sambasiva Rao, *J. Indian Council Chemists*, **2001**, 17, 33.
[37]. W. W. Hsieh, *Neural Networks*, **2000**, 13, 1105.
[38]. J. Takeuchi, Y. Kosugi, *Neural Networks*, **1994**, 7, 389.
[39]. N. Mai-Duy, T. Tran-Cong, *Neural Networks*, **2001**, 14, 185.
[40]. Y. Wang, C, Lin, IEEE Trans, *Neural Networks*, **1998**, 9, 294.
[41]. A. L. Giraud, J. C. Lemm, *Neural Networks*, **1995**, 8, 757.
[42]. J. Wang, *Neural Networks*, **1994**, 7, 629.
[43]. J. Molina-vilaplana, J. L. Pedreno-Molina, J. Lopez-Coronado, *Neurocomputing*, **2004**, 61, 495.
[44]. Y. Leung, K. Chen, Y. Jiao, X. Gao, S, Leung, IEEE Trans *Neural Networks*, **2001**, 12, 1074.
[45]. F. Friedrichs, M. Schmitt, *Neurocomputing*, **2005**, 63, 483.
[46]. M. S. Sanchez, L. A. Sarabia, *Chemom Intell Lab Syst*, **2002**, 63, 169.
[47]. M .V. Subba Rao, Ph. D thesis, Andhra University, Visakhapatnam, **2002**.
[48]. P. J. Gemperline, J. R. Long, G. Gregoriou, *Anal Chem*, **1991**, 63, 2313.
[49]. Z. Wang, J. Hwang, B, R, Kowalski, *Anal Chem*, **1995**, 67, 1487.
[50]. M. V. Subba Rao, V. Ananta Ramam, V. Muralidhara Rao, R. Sambasiva Rao, *Asian J. Chem*., **2010**, 22, 593.
[51]. I. Rojas, J. Gonzalez, A. Canas, A. F. Diaz, F. J. Rojas, M, Rodriguez, *Int. J Neural Syst*., **2000**, 10, 353.
[52]. R. J. Schlling, *IEEE Trans Neural Networks*, **2001**, 12, 1.
[53]. F. Scarselli, A. Tsir, *Neural Networks*, **1998**, 11, 35.
[54]. D. Stein, A. Feuer, *Neural Networks*, **1998,** 11, 235.
[55]. L. Jiao, *IEEE Trans Neural Networks*, **2001**, 12, 1060.
[56]. M. Burger, A. Nbauer, *Neural Networks*, **2003**, 16, 79.
[57]. J. Gonzalez, I. Rojas, H. Pomares, J. Ortega, *IEEE Trans Neural Networks*, **2002**, 13, 132.
[58]. J. M. Lee, Introduction to smooth manifolds. New York: Springer, **2012**.
[59]. F. H. Hamker, *Neural Networks*, **2001**, 14, 551.
[60]. H. Lipson, H. T. Siegelmann, *Neural Comput*., **2000**, 12, 2331.